Low-Rank Room Impulse Response Estimation

Martin Jälmby, Filip Elvander, and Toon van Waterschoot *Member, IEEE*

***Abstract*—In this paper we consider low-rank estimation of room impulse responses (RIRs). Inspired by a physics-driven room-acoustical model, we propose an estimator of RIRs that promotes a low-rank structure for a matricization, or reshaping, of the estimated RIR. This low-rank prior acts as a regularizer for the inverse problem of estimating an RIR from input-output observations, preventing overfitting and improving estimation accuracy. As directly enforcing a low rank of the estimate results is an NP-hard problem, we consider two different relaxations, one using the nuclear norm, and one using the recently introduced concept of quadratic envelopes. Both relaxations allow for implementing the proposed estimator using a first-order algorithm with convergence guarantees. When evaluated on both synthetic and recorded RIRs, it is shown that under noisy output conditions, or when the spectral excitation of the input signal is poor, the proposed estimator outperforms comparable existing methods. The performance of the two low-rank relaxations methods is similar, but the quadratic envelope has the benefit of superior robustness to the choice of regularization hyperparameter in the case when the signal-to-noise ratio is unknown. The performance of the proposed method is compared to that of ordinary least squares, Tikhonov least squares, as well as the Cramér-Rao lower bound (CRLB).**

***Index Terms*—Low-rank Modeling, Room Impulse Responses, Quadratic Envelopes**

I. Introduction

CONSIDERING the acoustics of a room as a linear time-invariant (LTI) system, room impulse responses (RIRs) describe the impact of the room on a sound signal, between a point source and a point receiver. They play a vital part in algorithms within a multitude of acoustic signal processing tasks, such as source localization [1], speech dereverberation [2], auralization [3], source separation [4], and echo cancellation [5].

There are several ways of modeling the RIR. Among the most popular ones are the infinite impulse response (IIR) (see e.g. [6], [7]) and finite impulse response (FIR) models (see e.g. [6], [8]). The IIR model offers the possibility of a more compact representation, however with the downside of possible difficulties estimating the filter parameters [9], and potential issues with instability [10]. The FIR model is simple and straightforward, but with the disadvantage that comparatively many coefficients are needed to accurately represent the RIR [9]. For a regular office-sized room, the FIR model can be several thousands of taps long [2], which can be prohibitive from a computational and memory requirement point for view [11], [12].

If the RIR is estimated from an output signal originating from an input signal with poor spectral excitation, i.e. lacking power in certain frequency bands, the estimation problem can become ill-posed, and the resulting estimation could suffer from large variance. To counteract this, it is common to use regularization, see e.g. [13]. The standard approach in RIR estimation is to use Tikhonov regularization [14], a special case of the types of regularization presented in [13]. However, we will demonstrate in this work that by exploiting an (approximate) low-rank structure of RIRs, estimation performance may be improved as compared to state-of-the-art estimators with and without regularization. As of late, low-rank modeling, and low-rank regularization, has become increasingly popular in signal processing, data science and related fields, such as machine learning [15], video background subtraction [16], and matrix completion [17].

In these applications, low-rank structures of matrices or tensors parametrizing the signals under consideration are exploited as to, e.g., obtain compact representations or to regularize otherwise ill-posed estimation problems. The class of signals amenable to this type of treatment includes polynomials, rational functions, smooth and periodic functions, and what will be used in this paper as a model for motivating low-rank structure of RIRs, sums of decaying sinusoids [18].

It should here be noted that in the system identification literature, low-rank structures are most commonly utilized for describing the Toeplitz matrix representing the discrete convolution operation, i.e., the LTI system acting on an input. This has been discussed by, e.g., Marconato *et al.* in [19] and [20]. In these works, empirical Bayes methods are primarily being used, and it is shown that low-rank regularization acts as a powerful method for improving the accuracy of the estimation. In particular, promoting Toeplitz matrices with low-rank is shown to add robustness in the face of input signals with poor spectral excitation.

In contrast, the low-rank structure considered herein does not refer to aforementioned Toeplitz matrix, but rather to a matricization of the impulse response itself. Specifically, we here aim to exploit that RIR vectors when reshaped into a matrix allow for low-rank approximations, something we have taken advantage of in our recent work [21] (see also [18], [22] for relevant work outside the domain of acoustic signal processing). System identification with this type of low-rank structure has earlier been explored by Paleologu *et al.* in works like [23]–[25] using iterative Wiener filter, recursive least-squares, and Kalman filter respectively, but there under the name *nearest Kronecker product*. The main focus there are shorter impulse responses, particularly for network echo cancellation. In contrast, the scenario considered in this paper

This research work was carried out at the ESAT Laboratory of KU Leuven. The research leading to these results has received funding from the Research Foundation – Flanders (FWO, grant 12ZD622N) and from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Manuscript received July 15, 2022; revised ...

concerns longer impulse responses for applications in acoustic signal processing.

Furthermore, the estimation algorithms in the above mentioned papers require the user to exactly specify the number of low-rank components to be estimated, thereby relying on oracle knowledge of the system complexity. In the work presented herein, we instead propose an estimator where the rank of the estimate is implicitly controlled by means of hyperparameters. In particular, the proposed estimator poses RIR estimation as an inverse problem, with low-rank estimates being promoted by the use of a regularization term.

It may here be noted that directly penalizing the rank of an RIR matrix yields an NP-hard problem [26]. A popular remedy to this in general low-rank estimation is nuclear norm regularization, i.e., penalizing the sum of the singular values, since it yields a convex optimization problem that can be solved by standard methods. This approach has however been shown to introduce a shrinkage, or bias towards zero, of the singular values (see e.g. [15]).

In this work, we propose to counter this bias by utilizing the so-called quadratic envelope of the rank penalty. Quadratic envelopes, introduced in [27], constitute a class of minorizers constructed from point-wise best function approximations from below by quadratic functions, and have recently been used for addressing bias issues in sparsity problems within the realm of image processing and computer vision by Carlsson *et al.* in works like [27]–[30]. A drawback of this relaxation is that the resulting optimization problem is not necessarily convex, but in many cases this is outweighed by the very attractive property that, for appropriate hyperparameter values, the relaxation does not move any local minima of the problem being approximated. Its merit is, however, to the best of the authors' knowledge, yet to be explored in the context of acoustic signal processing. In addition, the problem considered herein falls into a class of problems for which convergence of the proximal gradient method, or forward-backward splitting (FBS), to a local minimum can be guaranteed.

The contribution of this paper is therefore to provide a framework for low-rank estimation of room impulse responses. We present an easy-to-use algorithm and show that low-rank regularization outperforms benchmark methods, that is, ordinary least squares and Tikhonov regularized least squares. This is demonstrated using synthetic as well as real-life RIRs, driven by synthetic as well as real-life signals and it is shown that this works well on real life-recorded RIRs. The outperformance is consistent, but particularly noticeable when the SNR is low, or when the spectral excitation is poor. Further, the quadratic envelope performs exceptionally well when prior knowledge about the SNR can not be assumed. Finally, it should be noted that the use of the proposed algorithm is not restricted to acoustics and RIRs, but rather it is applicable to any physical system of which the impulse response is well modeled by a sum of decaying sinusoids.

This paper is organized as follows: first, Section I is concluded with an introduction of the notation used throughout the paper. In Section II, the signal model is introduced. In Section III, different possible relaxations of the low-rank penalty term are discussed, and in Section IV, the proposed algorithm is introduced. Numerical results are presented in Section V. Finally, in Section VI, conclusions of the work presented here are discussed, and possible avenues for future research are being pointed out.

A. Notation

Vectors are denoted by lower case, bold letters, such as $\mathbf{h}$, and matrices are denoted by upper case, bold letters, such as $\mathbf{H}$. Subscript on an upper case bold letter indicates matrix column, i.e. $\mathbf{H}_j$ denotes the jth column of $\mathbf{H}$, and on a lower case bold letter it indicates vector element, i.e. $\mathbf{h}_j$ denotes the jth element of $\mathbf{h}$. A bar, $\bar{\cdot}$, denotes complex conjugation of a complex number and the hat symbol, $\hat{\cdot}$, denotes an estimated quantity. The symbol $\odot$ denotes the Hadamard product, i.e. element-wise multiplication of vectors or matrices, the nabla operator, ∇, refers to the gradient of a function, and $\mathbb{E}[\cdot]$ denotes statistical expectation. Linear operators are denoted by upper case calligraphic letters, such as $\mathcal{A}$, and an asterisk, $\cdot^*$, denotes adjoint. Sets are also denoted by upper case calligraphic letters, but it will be obvious from context what is considered. The norm of a matrix or an operator is denoted $\|\cdot\|$ and refers to $\|\cdot\|_2$, $\langle \mathbf{u}, \mathbf{v} \rangle_n = \sum_{j=1}^{n} \mathbf{u}_j \mathbf{v}_j$ denotes the inner product between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and $i = \sqrt{-1}$ denotes the imaginary unit.

II. Signal Model

We assume an observed acoustic pressure signal in discrete time $y(n), n = 1, 2, \ldots, n_y$, with corresponding signal vector $\mathbf{y} \in \mathbb{R}^{n_y}$. This observed signal is the result of a known signal $x(n), n = 1, 2, \ldots, n_x$, with corresponding vector $\mathbf{x} \in \mathbb{R}^{n_x}$, being generated by a point source in a room, and can therefore be modeled as the convolution of $\mathbf{x}$ with the room impulse response $h(n), n = 1, 2, \ldots, n_h$, with corresponding vector $\mathbf{h} \in \mathbb{R}^{n_h}$, $\mathbf{x} * \mathbf{h}$, where

$$(\mathbf{x} * \mathbf{h})_n = \sum_{k=1}^{n_h} \mathbf{h}_k \mathbf{x}_{n+n_h-k}, \tag{1}$$

for $n = 1, 2, \ldots, n_y$, i.e. we consider only the part of the RIR where the system is fully excited. We will assume that $n_x \geq n_h$, to not have to consider initial conditions. We further assume measurement noise $\mathbf{e} \in \mathbb{R}^{n_y}$, that we assume to be white Gaussian with variance σ^2, i.e. $\mathbf{e} \in \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_y})$, where $\mathbf{I}_{n_y} \in \mathbb{R}^{n_y \times n_y}$ denotes the identity matrix. In the more general case of non-white, non-stationary noise, a combination of prefiltering and regularization can be used, as elaborated on in [14]. Taken together, we have that

$$\mathbf{y} = \mathbf{x} * \mathbf{h} + \mathbf{e}, \tag{2}$$

which can also be written as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}, \tag{3}$$

where $\mathbf{X} \in \mathbb{R}^{n_y \times n_h}$ is the Toeplitz matrix corresponding to the convolution operation in (2).

We will have that $n_y = n_x - n_h + 1$. We define the linear operator $\mathcal{A} : \mathbb{R}^{n_h} \to \mathbb{R}^{n_y}$, such that $\mathcal{A}(\mathbf{h}) = \mathbf{x} * \mathbf{h}$. Here we also define the adjoint operator $\mathcal{A}^* : \mathbb{R}^{n_y} \to \mathbb{R}^{n_h}$, as the unique operator for which $\langle \mathbf{h}, \mathcal{A}^*(\mathbf{y}) \rangle_{n_h} = \langle \mathcal{A}(\mathbf{h}), \mathbf{y} \rangle_{n_y}, \forall \mathbf{h} \in \mathbb{R}^{n_h}, \mathbf{y} \in \mathbb{R}^{n_y}$.

III. Estimation with Low-rank Heuristics

The room impulse response $\mathbf{h}$ can be estimated from the input and observed signal vectors with the least squares method,

$$\underset{\mathbf{h}}{\text{minimize}}\, \frac{1}{2}\|\mathbf{y} - \mathcal{A}(\mathbf{h})\|_2^2, \tag{4}$$

and under the assumption that the noise $\mathbf{e}$ is white Gaussian, this yields the maximum likelihood estimator. Inverse problems like these are, however, often ill-posed [31], [32], particularly in acoustic signal processing, where poor excitation is a common occurrence [14]. The inverse mapping from $\mathbf{y}$ to $\mathbf{h}$ might therefore be unstable due to ill-conditioning. The fact that the output is corrupted by noise could yield an over-fitted solution, and finally, the solution could be non-unique [32]. In order to have a well-posed problem, a possible approach is to use regularization [13]. This is done by appending a penalty term to the optimization problem (4). With that, the problem can be written as

$$\underset{\mathbf{h}}{\text{minimize}}\, f(\mathbf{h}) + g(\mathbf{h}). \tag{5}$$

We consider the function $f : \mathbb{R}^{n_h} \to \mathbb{R}$ to be the data-fit term $\frac{1}{2}\|\mathbf{y} - \mathcal{A}(\mathbf{h})\|_2^2$, i.e. a function penalizing the distance between the model output and the observation. The function $g : \mathbb{R}^{n_h} \to \mathbb{R}$ is the regularization term. Note that this seemingly simple form includes constrained optimization problems, since we can let g be an indicator function for some non-empty set $\mathcal{S}$,

$$g(\mathbf{h}) = \mathcal{I}_{\mathcal{S}}(\mathbf{h}) = \begin{cases} 0 & , \mathbf{h} \in \mathcal{S} \\ \infty & , \mathbf{h} \notin \mathcal{S} \end{cases}. \tag{6}$$

This form also includes the aforementioned Tikhonov regularization, where $g(\mathbf{h}) = \lambda_{\text{T}}\|\mathbf{h}\|_2^2$, yielding

$$\underset{\mathbf{h}}{\text{minimize}}\, \frac{1}{2}\|\mathbf{y} - \mathcal{A}(\mathbf{h})\|_2^2 + \lambda_{\text{T}}\|\mathbf{h}\|_2^2, \tag{7}$$

where $\lambda_{\text{T}} \geq 0$ is a parameter controlling the degree of regularization.

The regularization should, however, preferably be designed using prior knowledge about the problem [14]. We will therefore demonstrate the low-rank properties of RIRs, and argue how these can be exploited in regularization. An RIR can be well modelled as an infinite sum of decaying sinusoids, see e.g. [33]. The frequencies of these decaying sinusoids correspond to the modes of the room, i.e. frequencies at which standing waves would occur in the absence of wall absorption. The number of room modes, N_f, below a certain frequency f satisfies $N_f \approx \frac{4\pi}{3c^3} f^3$ [34], where c is the speed of sound. One might therefore expect to have to use a large number of terms to approximate this sum. However, in [10] it is concluded that the large overlap between the modal curves, which occurs particularly at higher frequencies [9], makes the number of distinguishable peaks in the frequency magnitude response much smaller, and it is justified to approximate the infinite sum with a finite one. For a longer exposition, see e.g. [10] and [21]. With this, the room impulse response can be modeled by

$$h(\mathbf{r}_r, \mathbf{r}_s, n) = \sum_{m=1}^{m_s} \mu_m(\mathbf{r}_r, \mathbf{r}_s) e^{-\beta_m n} \cos(\omega_m n + \phi_m), \tag{8}$$

for $n = 1, 2, \ldots, n_h$. Here, μ_m denotes the initial amplitude, $\mathbf{r}_r, \mathbf{r}_s \in \mathbb{R}^3$ the position of the receiver and the source, respectively, $\beta_m \in \mathbb{R}_+$ the exponential decay constant, $\omega_m \in [0, \pi]$ the angular frequency $\phi_m \in [0, 2\pi)$, the phase, and $m_s \in \mathbb{N}$ is the number of decaying sinusoids used in the model. For ease of notation we will drop the dependence on $\mathbf{r}_r$ and $\mathbf{r}_s$ and refer to $h(\mathbf{r}_r, \mathbf{r}_s, n)$ simply as $h(n)$, the same $h(n)$ as presented in Section II.

When a vector consisting of the sum of m_s discrete time decaying sinusoids is reshaped into a matrix, that matrix will have rank $2m_s$, see e.g. [35]. Further, as previously mentioned, there is a large spectral overlap of some of the modal curves of the decaying sinusoids that make up the RIR, particularly at higher frequencies. These two facts taken together makes for good conditions for enforcing a low-rank structure on the solution, i.e. that when the solution vector is reshaped into a matrix, that matrix should have low rank. In previous work we have shown the connection between the physics driven modeling of room acoustics and the use of the sum of decaying sinusoids model, along with the connection to low rank modeling [21]. For more on the physical motivation we refer to [21], but for the convenience of the reader, we here reiterate the low-rank motivation. It should be noted that a measured RIR will, when matricized, not be low-rank in the strict sense, because of measurement noise and model errors. *Low-rank* is here meant in a less strict sense, i.e. that a low-rank approximation of the RIR will render a small approximation error.

Let us define the reshaping operator $\mathcal{R}_{pq} : \mathbb{R}^{n_h} \to \mathbb{R}^{p \times q}$, that reshapes a vector $\mathbf{h} \in \mathbb{R}^{n_h}$ into a matrix $\mathbf{H} \in \mathbb{R}^{p \times q}$, for which $n_h = pq$. Then, if we assume that $q = \sqrt{n_h} \in \mathbb{N}$, the RIR vector $\mathbf{h}$ can be reshaped into a square matrix $\mathbf{H} = \mathcal{R}_{qq}(\mathbf{h}) \in \mathbb{R}^{q \times q}$,

$$\mathbf{H} = \begin{bmatrix} h(1) & h(q+1) & \ldots & h(q(q-1)+1) \\ \vdots & \vdots & & \vdots \\ h(q) & h(2q) & \ldots & h(n_h) \end{bmatrix}. \tag{9}$$

Then, $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2 + \cdots + \mathbf{H}_{m_s}$, where $\mathbf{H}_m$ corresponds to the mth decaying sinusoid of (8),

$$\begin{aligned} \mathbf{H}_m &= \begin{bmatrix} h_m(1) & h_m(q+1) & \ldots & h_m(q(q-1)+1) \\ \vdots & \vdots & & \vdots \\ h_m(q) & h_m(2q) & \ldots & h_m(n_h) \end{bmatrix} \\ &= \frac{\mu_m e^{i\phi_m}}{2} \begin{bmatrix} z_m \\ z_m^2 \\ \vdots \\ z_m^q \end{bmatrix} \begin{bmatrix} 1 & z_m^q & \cdots & z_m^{(q-1)q} \end{bmatrix} \\ &\quad + \frac{\mu_m e^{-i\phi_m}}{2} \begin{bmatrix} \bar{z}_m \\ \bar{z}_m^2 \\ \vdots \\ \bar{z}_m^q \end{bmatrix} \begin{bmatrix} 1 & \bar{z}_m^q & \cdots & \bar{z}_m^{(q-1)q} \end{bmatrix}, \end{aligned} \tag{10}$$

where $z_m = e^{i\omega_m - \beta_m}$. We see that each of the m_s terms that makes up $\mathbf{H}$, can be written as a sum of two rank-1 matrices, i.e. as a rank-2 matrix. As long as z_m are distinct, the matrix $\mathbf{H}$ will have rank $2m_s$ [36]. In general, $\mathbf{H}$ could be a non-square

matrix, but here we consider the square case, as to simplify the exposition.

In light of this discussion, the problem to be solved would ideally be

$$\begin{aligned} \underset{\mathbf{h}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{y}-\mathcal{A}(\mathbf{h})\|_2^2 \\ \text{s.t.} \quad & \text{rank}(\mathcal{R}_{qq}(\mathbf{h})) \leq 2m_s, \end{aligned} \tag{11}$$

or, equivalently, using the form of (5),

$$\underset{\mathbf{h}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y}-\mathcal{A}(\mathbf{h})\|_2^2 + \mathcal{I}_{2m_s}\left(\mathcal{R}_{qq}(\mathbf{h})\right), \tag{12}$$

where the penalty function, $\mathcal{I}_{2m_s}$, is the indicator function for the set of matrices of rank at most $2m_s$. This problem, however, is non-convex and NP-hard [26]. Furthermore, the exact rank, $2m_s$, is typically not known. Instead of the constrained problem of (11) and (12), one could consider the penalized version,

$$\underset{\mathbf{h}}{\text{minimize}} \frac{1}{2}\|\mathbf{y}-\mathcal{A}(\mathbf{h})\|_2^2 + \lambda\,\text{rank}(\mathcal{R}_{qq}(\mathbf{h})). \tag{13}$$

Here, $\lambda \geq 0$ is a parameter controlling the degree of penalization of the rank of the solution, in analogy to λ_{T} in (7). This transforms the problem of not knowing the rank $2m_s$, into the problem of choosing λ, but the problem remains NP-hard [15]. Considering that the rank of a matrix is equal to the number of non-zero singular values, (13) is equivalent to

$$\underset{\mathbf{h}}{\text{minimize}} \frac{1}{2}\|\mathbf{y}-\mathcal{A}(\mathbf{h})\|_2^2 + \lambda\|\sigma(\mathcal{R}_{qq}(\mathbf{h}))\|_0. \tag{14}$$

where $\|\cdot\|_0$ is the ℓ_0-pseudo-norm and $\sigma(\mathbf{H}) : \mathbb{R}^{q\times q} \to \mathbb{R}^q$ is the function that takes an $\mathbb{R}^{q\times q}$-matrix and returns a vector of its singular values in non-decreasing order, given its singular value decomposition (SVD) $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, for $\mathbf{U} \in \mathbb{R}^{q\times q}$, $\mathbf{S} \in \mathbb{R}^{q\times q}$, $\mathbf{V}^T \in \mathbb{R}^{q\times q}$. The NP-hardness persists, however, so in order to be able to find an approximate solution to (14), the problem needs to be relaxed. This can be done in different ways, which will now be further explored.

A. Nuclear Norm Regularization

In order to achieve low-rank solutions, throughout this paper, two different kinds of penalty functions will be used and compared. Firstly, the popular approach of using the ℓ_1-norm instead of the ℓ_0-norm in the penalization of $\sigma(\mathbf{H})$ [37],

$$\underset{\mathbf{h}}{\text{minimize}} \frac{1}{2}\|\mathbf{y}-\mathcal{A}(\mathbf{h})\|_2^2 + \lambda_{\text{NN}}\|\mathcal{R}_{qq}(\mathbf{h})\|_*, \tag{15}$$

where $\|\mathbf{H}\|_*$ denotes the nuclear norm of $\mathbf{H}$, defined as $\|\mathbf{H}\|_* = \text{trace}((\mathbf{H}^*\mathbf{H})^{1/2}) = \sum_{n=1}^{q}\sigma_n(\mathbf{H}) = \|\sigma(\mathbf{H})\|_1$, and $\lambda_{\text{NN}} \geq 0$ is analogous to λ in (13). This is the first of the two optimization problems we will aim to solve in Section IV. The advantage of problem (15) is that it is convex and can be solved efficiently [38]. The disadvantage is that this solution will be biased, in that the singular values will be shrunk [15].

Several different methods have been proposed in an attempt to remedy the issue of the shrinking of large singular values. Among them, the weighted nuclear norm [39], which has been generalized to so called *generalized singular value thresholding* in [40].

B. Quadratic Envelope

The second approach to achieving a low-rank solution that will be used in this paper, is the use of a function created in attempt to closer emulate the ℓ_0-pseudonorm, the so called *quadratic envelope* of the ℓ_0-norm. The quadratic envelope of a function consists of pointwise, quadratic approximations, that nowhere overestimate the original function. For the function $g : \mathbb{R}^n \to \mathbb{R}$, the quadratic envelope, at the point $\mathbf{u} \in \mathbb{R}^n$, is defined as

$$\begin{aligned} &\mathcal{Q}_\gamma(g)(\mathbf{u}) = \sup_{\alpha\in\mathbb{R},\mathbf{v}\in\mathbb{R}^n} \\ &\left\{\alpha - \frac{\gamma}{2}\|\mathbf{u}-\mathbf{v}\|^2 : \alpha - \frac{\gamma}{2}\|\mathbf{z}-\mathbf{v}\|^2 \leq g(\mathbf{z}),\ \forall\,\mathbf{z}\right\}, \end{aligned} \tag{16}$$

i.e. the supremum of all functions of the form $\alpha - \frac{\gamma}{2}\|\mathbf{u}-\mathbf{v}\|^2$, that nowhere overestimates the original function $g(\cdot)$. The parameter $\gamma > 0$ controls how close this envelope will be to the original function. A smaller value of γ means a larger degree of smoothing of the function, whereas $\mathcal{Q}_\gamma(g)$ converges pointwise to g as $\gamma \to \infty$ [27].

Firstly, it is readily verified that

$$\lambda\|\mathbf{u}\|_0 = \lambda\sum_{n=1}^{n_u}\|u_n\|_0, \tag{17}$$

In other words, the problem decouples in all dimensions, and we can, without loss of generality, look at a one-dimensional version of the aforementioned function, i.e. $\mathbf{u} \equiv u$. This has been considered before, in works like [41] and [42], with support of the theory in [43] and [44], but we restate the results here. Further, since this is to be applied to the singular values, we will, to simplify the exposition, only consider $u \geq 0$.

Proposition: The quadratic envelope of $\lambda_{\text{QE}}\|u\|_0$ is given by

$$\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(u) = \begin{cases} \sqrt{2\lambda_{\text{QE}}\gamma}u - \frac{\gamma}{2}u^2 & ,\ u < \mu \\ \lambda_{\text{QE}} & ,\ \mu \leq u \end{cases} \tag{18}$$

or, equivalently,

$$\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(u) = \lambda_{\text{QE}} - \frac{\gamma}{2}\text{max}\left(\mu-u,0\right)^2, \tag{19}$$

where $\mu = \sqrt{2\lambda_{\text{QE}}/\gamma}$, and $\lambda_{\text{QE}} \geq 0$ is analogous to λ_{NN} in (15).

Proof: The proof is based on [44], but for the convenience of the reader we iterate it here. See Appendix A.

The major benefit of the quadratic envelope, as compared to the nuclear norm, manifests itself in (18). For values u larger than μ, the quadratic envelope is constant. The differences between the nuclear norm and the quadratic envelope in approximating $\|u\|_0$ are further illustrated in Fig. 1. The effect of the parameter γ is noticeable, in that a larger γ yields an envelope closer to $\|u\|_0$. Further, it can be seen that the main effect of increasing λ_{QE} is raising the level at which the function remains constant.

A particular case of the penalty (19), when $\gamma = 2$, has been considered in [29]. Adding (19) to (4) yields the final form of the second optimization problem we are aiming to solve in Section IV,

$$\underset{\mathbf{h}}{\text{minimize}} \frac{1}{2}\|\mathbf{y}-\mathcal{A}(\mathbf{h})\|_2^2 + \mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)\left(\mathcal{R}_{qq}\left(\mathbf{h}\right)\right), \tag{20}$$

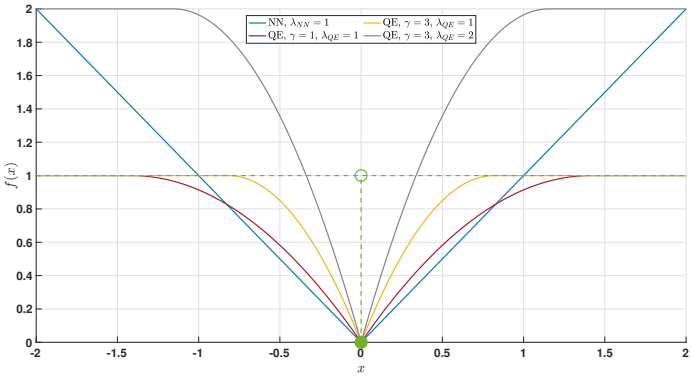

Fig. 1. Nuclear norm and quadratic envelope, for varying parameter values. For reference, the ℓ_0-pseudo-norm is indicated by the dashed green line.

where we, for ease of notation, for a matrix $\mathbf{X}$, define $\mathcal{Q}(\mathbf{X}) = \mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(\mathbf{X})$.

As previously mentioned, γ is a parameter that controls the fidelity of $\mathcal{Q}_\gamma(g)$ to the original function g. If $\gamma \leq \|\mathcal{A}\|$, then the objective function in (20) is convex. This is of course a very desirable property as these problems are generally considered comparatively easy to solve. However, for $\gamma > \|\mathcal{A}\|$ we have the following upside: if $\hat{\mathbf{h}}$ is a (strict) local minimizer of (20), then $\hat{\mathbf{h}}$ is also a (strict) local minimizer of (14). In addition to this, their global minima coincide [27].

IV. Algorithm

A. Preliminaries

As previously indicated, the optimization problem will be solved using the proximal gradient method, also known as forward-backward splitting (FBS), which can be seen in Algorithm 1. It has become very popular recently due to its ability to handle non-smooth functions. For more on the proximal gradient method, see e.g. [32], [45], [46]. The general idea is to alternate between taking a step in the negative direction of the gradient of the smooth data-fit term f, and trying to minimize the (possibly non-smooth) penalty function g, using the so-called *proximal operator*. For a function $g(\mathbf{V}) : \mathbb{R}^{q\times p} \to \mathbb{R}$, with the step length parameter $\rho > 0$, it is defined as

$$\text{prox}_{\rho g}(\mathbf{V}) = \arg\min_{\mathbf{Z}} \left(g(\mathbf{Z}) + \frac{1}{2\rho}\|\mathbf{Z}-\mathbf{V}\|_F^2 \right). \tag{21}$$

In Algorithm 1, the gradient step, $\mathbf{u}^k - \rho\nabla f(\mathbf{u}^k)$, is to be interpreted as the forward step, and the proximal mapping, $\text{prox}_{\rho g}(\cdot)$ as the backward step.

Algorithm 1: Proximal Gradient Method (PG)

Set $\mathbf{u}^0 \in \mathbb{R}^n$ and $\rho \in (0, 2L^{-1}]$
for $k = 0, 1, \ldots$ **do**
  $\mathbf{u}^{k+1} = \text{prox}_{\rho g}(\mathbf{u}^k - \rho\nabla f(\mathbf{u}^k))$

Here, L denotes the Lipschitz constant of ∇f. In order to guarantee that Algorithm 1 will converge, a few additional technical assumptions have to be made about f and g respectively. These are all met for the problems considered in this paper, more details can be found in [47].

In order to facilitate the exposition we will introduce a couple of linear operators. The operator $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^n$ denotes the reversion of the order of the elements in a vector, i.e $\mathcal{H}(\mathbf{x}) = [\mathbf{x}_{n_x}, \mathbf{x}_{n_x-1}, \ldots, \mathbf{x}_1]^T$. The operator $\mathcal{F} : \mathbb{R}^n \to \mathbb{C}^n$ denotes the discrete Fourier transform (DFT), and $\mathcal{F}^{-1} : \mathbb{C}^n \to \mathbb{R}^{n1}$ the inverse DFT. The operator $\mathcal{P}^k : \mathbb{R}^n \to \mathbb{R}^{n+k}$ denotes padding of a vector with k zeroes at the end. Finally, the operator $\mathcal{C}_k^n : \mathbb{R}^m \to \mathbb{R}^{n-k+1}$, where $n \geq k$ and $m \geq n-k+1$, is the linear operator that picks out the kth till the nth element of a vector, i.e. $\mathcal{C}_k^n(\mathbf{y}) = \begin{bmatrix}\mathbf{y}_k, & \mathbf{y}_{k+1}, & \ldots, & \mathbf{y}_n\end{bmatrix}$.

B. Proposed Algorithm

The previously introduced operators $\mathcal{A}$ and $\mathcal{A}^*$ can now be expressed in terms of these operators,

$$\mathcal{A}(\mathbf{h}) = \mathcal{C}_{n_h}^{n_x}\left(\mathcal{F}^{-1}\left(\mathcal{F}\left(\mathcal{P}^{n_x-1}(\mathbf{h})\right) \odot \mathcal{F}\left(\mathcal{P}^{n_h-1}(\mathbf{x})\right)\right)\right), \tag{22}$$

and

$$\mathcal{A}^*(\mathbf{y}) = \mathcal{C}_{n_y}^{n_x}\left(\mathcal{F}^{-1}\left(\mathcal{F}\left(\mathcal{P}^{n_x-n_h}(\mathcal{H}(\mathbf{x}))\right) \odot \mathcal{F}\left(\mathcal{P}^{2n_x-n_h-n_y}(\mathbf{y})\right)\right)\right), \tag{23}$$

respectively.

The first step in iteratively solving (15) and (20), respectively, is the gradient step, with respect to the data-fit term f in the cost function. The gradient of (4) is given by

$$\nabla f = \mathcal{A}^*(\mathcal{A}(\mathbf{h})) - \mathcal{A}^*(\mathbf{y}). \tag{24}$$

For the backward step, there is a seemingly complicating fact in that this is to be taken with respect to a matrix and its singular values. This, however, turns out not to be a problem. After the gradient step, the vector is reshaped into a matrix, and an SVD is performed on that matrix.

Proposition: The proximal operator of $\mathcal{Q}(\mathbf{H})$ is given by

$$\text{prox}_{\rho\mathcal{Q}(\cdot)}(\mathbf{H}) = \mathbf{U}\,\mathcal{D}\left(\text{prox}_{\rho\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)}\left(\sigma(\mathbf{H})\right)\right)\mathbf{V}^T, \tag{25}$$

and the proximal operator of $\lambda_{\text{NN}}\|\sigma(\mathbf{H})\|_1$ is given by

$$\text{prox}_{\rho\lambda_{\text{NN}}\|\cdot\|_1}(\mathbf{H}) = \mathbf{U}\,\mathcal{D}\left(\text{prox}_{\rho\lambda_{\text{NN}}\|\cdot\|_1}\left(\sigma(\mathbf{H})\right)\right)\mathbf{V}^T, \tag{26}$$

where $\mathcal{D} : \mathbb{R}^q \to \mathbb{R}^{q\times q}$ is the operation of creating a diagonal $q\times q$-matrix, with the argument of the operator as the diagonal.

Proof: See Appendix B. It may be noted that this is due to the orthogonal invariance of $\mathcal{Q}(\mathbf{H})$ and $\lambda_{\text{NN}}\|\sigma(\mathbf{H})\|_1$, respectively [46].

After either of the proximal operators has acted on the vector of singular values, all that remains is to reassemble the matrix, with the adjusted singular values, and then vectorize, in order to get the current iterate of the estimated RIR. Finally, a check is made to see if the algorithm is making enough progress, i.e. if the new iterate is sufficiently different from the previous one, or if the iterative procedure should be terminated. These steps are summarized in Algorithm 2.

[1]In the general case, $\mathcal{F}^{-1} : \mathbb{C}^n \to \mathbb{C}^n$, but because of the conjugate symmetry properties of the signals considered here, we will have $\mathcal{F}^{-1} : \mathbb{C}^n \to \mathbb{R}^n$

Solving (15) or (20) differs only in the use of proximal operator. That is reflected in step 4 of Algorithm 2, in that 4 a) corresponds to solving (15) and 4 b) corresponds to solving (20).

Algorithm 2: Low-Rank RIR Estimation Algorithm

Input: $\mathbf{h}^0$, ρ, γ, λ, $\mathbf{x}$, $\mathbf{y}$, tol, maxIter
for $k = 0, 1, \ldots,$ maxIter **do**
1: $\mathbf{w}^k = \mathbf{h}^k - \rho(\mathcal{A}^*(\mathcal{A}(\mathbf{h}^k)) - \mathcal{A}^*(\mathbf{y}))$ {Forward step}
2: $\mathbf{W}^k = \mathcal{R}_{qq}(\mathbf{w}^k)$
3: $[\mathbf{U}^k, \mathbf{S}^k, \mathbf{V}^k] = \text{SVD}(\mathbf{W}^k)$
4 a): $\mathbf{s}^{k+1} = \mathcal{S}_{\lambda_{\text{NN}}}\left(\sigma(\mathbf{W}^k)\right)$
4 b): $\mathbf{s}^{k+1} = \text{prox}_{\rho\mathcal{Q}(\cdot)}(\mathbf{W}^k)$
5: $\mathbf{H}^{k+1} = \mathbf{U}^k\mathcal{D}(\mathbf{s}^{k+1})\mathbf{V}^{k^T}$ {Backward step}
6: $\mathbf{h}^{k+1} = \mathcal{R}_{qq}^{-1}(\mathbf{H}^{k+1})$
if $\|\mathbf{h}^{k+1} - \mathbf{h}^k\|/(\|\mathbf{h}^k\| + \delta) \leq$ tol **then**
break
end if
end for

The parameter ρ in (21) controls the balance between minimizing the penalty function, and staying close the current point $\mathbf{V}$, and serves as step-size for the proximal gradient algorithm. Small values of ρ will yield a $\mathbf{Z}$ close to $\mathbf{V}$, whereas larger values of ρ will result in $\mathbf{Z}$ being closer to the minimum of $g(\mathbf{Z})$. In order to assure that the sequence generated by Algorithm 1 is bounded, we must chose $\rho \in (0, 2/L)$ [48].

We have that $\|\nabla f(\mathbf{w}) - \nabla f(\mathbf{z})\| \leq \|\mathcal{A}^*\mathcal{A}\|\|\mathbf{w} - \mathbf{z}\|$ i.e. $L \leq \|\mathcal{A}^*\mathcal{A}\|$. For a linear operator, the norm of the operator is the same as for the corresponding matrix, i.e. $\|\mathcal{A}^*\mathcal{A}\| = \|\mathbf{X}^T\mathbf{X}\|$, where $\mathbf{X} \in \mathbb{R}^{n_y \times n_h}$ is the Toeplitz matrix s.t. $\mathcal{A}(\mathbf{h}) = \mathbf{X}\mathbf{h}$. Computing $\|\mathbf{X}^T\mathbf{X}\|$ is, however, computationally demanding and something that should be avoided if possible. In this instance, it is possible to use a bound on $\|\mathbf{X}\|$. It is shown in [49] that $\|\mathbf{X}\| \leq 2M$, where M denotes the essential supremum of the absolute value of the Fourier series of $\mathbf{x}$. Taken together with the well-established fact that, for any two linear operators $\mathcal{A}$ and $\mathcal{B}$, $\|\mathcal{A}\mathcal{B}\| \leq \|\mathcal{A}\|\|\mathcal{B}\|$, we have that $L \leq 4M^2$. In order to ensure that the strict inequality is upheld, we let

$$\rho = 0.99\left(2M^2\right)^{-1}. \tag{27}$$

Finally, three more parameters need to be set by the user. A maximum number of iterations, *maxIter*, that the user is willing to run, a tolerance level, *tol*, to determine when the algorithm is not making sufficient progress anymore and should be terminated, and a small number δ to avoid division by zero in the normalization of the norm of the difference between the new and the old iterate. Preliminary simulations showed that *maxIter* $= 10^4$, *tol* $= 10^{-4}$, and $\delta = 10^{-6}$ were suitable choices, and these values will be used throughout this paper.

C. Computational Complexity

In this Section we study the computational complexity of Algorithm 2, as a function of the length of the RIR, n_h. The difference between using the nuclear norm or quadratic envelope penalty is negligible, in terms of complexity per iteration. Asymptotically, as $n_h \to \infty$, the most demanding steps of the algorithm are 3 and 5. In step 3, we perform a singular value decomposition of the $\sqrt{n_h} \times \sqrt{n_h}$-matrix $\mathbf{W}^k$, which requires $\mathcal{O}(n_h^{3/2})$ multiplications. Similarly, the two $\sqrt{n_h} \times \sqrt{n_h}$-matrix multiplications in step 5 require, if naively executed, $\mathcal{O}(n_h^{3/2})$ multiplications.

In terms of wall time, for problems of the size studied in this paper, step 1 of the algorithm is the most demanding. Although the fast Fourier transform (FFT) require only $\mathcal{O}(n_h \log(n_h))$ multiplications, the fact that we in step 1 have to carry out 4 of these (including the IFFT's) each iteration, makes this the most time-consuming part of the algorithm.

V. Numerical Results

In order to visualize the dependence of the proposed and benchmark algorithms' performance on room related parameters, we will first provide numerical results using synthetically generated RIRs. Then, to demonstrate the practical applicability of the proposed algorithm, we proceed by presenting results of simulations using real-life RIRs. The accuracy of an RIR estimate, $\hat{\mathbf{h}}$, is measured by the *normalized misalignment*, defined as

$$\mathcal{M}_{\text{dB}}\left(\hat{\mathbf{h}}\right) = 20\log_{10}\left(\frac{\|\hat{\mathbf{h}} - \mathbf{h}\|_2}{\|\mathbf{h}\|_2}\right), \tag{28}$$

where $\mathbf{h}$ is the RIR we are aiming to estimate. In Sections V-I and V-J we will use slightly modified evaluation measures, which will be further explained there. Throughout this Section we will compare the performance of the two proposed methods to that of two benchmark methods. Firstly, ordinary least squares, i.e. without regularization, corresponding to solving (4). Secondly, least squares, with Tikhonov regularization, i.e. solving

$$\underset{\mathbf{h}}{\text{minimize}}\, \frac{1}{2}\|\mathbf{y} - \mathcal{A}(\mathbf{h})\|_2^2 + \lambda_{\text{T}}\|\mathbf{h}\|_2^2, \tag{29}$$

where λ_{T} controls the degree of regularization.

A. Impact of T_{60}

First we present simulations performed on synthetically generated RIRs, in order to show how the algorithms under comparison perform, as a function of the reverberation time, T_{60}, of the room. The T_{60} is defined as the time required for the sound level to drop 60 dB after switching off a stationary source. The RIRs are generated by letting each tap independently be drawn from a normal distribution, with an exponentially decaying envelope, corresponding to the given T_{60}. This is a stochastic reverberation model [50], and can be attributed to [51] and [52]. In this simulation example, for each value of T_{60}, taken from a a grid of 9 values in the range [33, 300] ms, 10 synthetic RIRs were generated. The system was driven by a white Gaussian noise. Finally, independent white Gaussian measurement noise was added, with signal-to-noise ratio (SNR_{dB}) $= 10$, defined as

$$\text{SNR}_{\text{dB}} = 10\log_{10}\left(\frac{P_S}{P_N}\right), \tag{30}$$

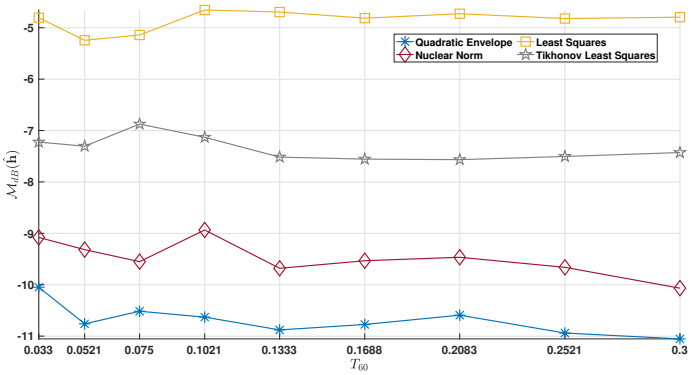

Fig. 2. Normalized misalignment as a function of T_{60}

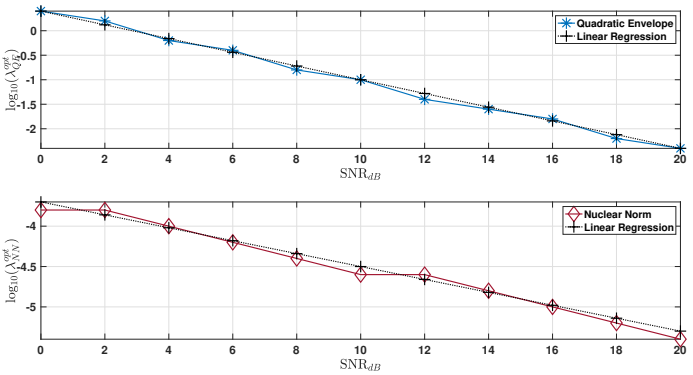

Fig. 3. Linear regression for λ_{QE} (top) and λ_{NN} (bottom), respectively, as a function of $\mathrm{SNR}_{\mathrm{dB}}$.

where P_S and P_N denote the power of the output signal without the noise, and the power of the noise, respectively. The RIR length, n_h, was set to correspond to the reverberation time, and n_y was set to $1.2n_h$, in order to have comparable overdetermined systems, for the various values of T_{60}. The hyperparameters λ_{T}, λ_{NN}, λ_{QE}, and γ were tuned for each individual setting, by finding the optimal values using another set of randomly generated RIRs. The results are shown in Fig. 2. There it can be seen that the performance is close to constant as a function of T_{60} for all the analyzed methods, and that the low-rank models outperform ordinary least squares, as well as Tikhonov-regularized least squares, with a slight preference for the quadratic envelope.

B. SMARD

In order to demonstrate the proposed methods' applicability to actual measured RIRs, we apply it to the *single- and multichannel audio recordings database* (SMARD) [53]. These recordings are sampled at 48 kHz, for 12 seconds, yielding impulse responses of $576 \cdot 10^3$ taps. These are recorded at various source and receiver positions, and with varying equipment, in a room of size 7.34 m $\times 8.09$ m $\times 2.87$ m, with a reverberation time of approximately 0.15 s. In total, the dataset contains 1008 RIRs. Two things should be noted regarding the numerical simulations performed here. Firstly, the RIRs are truncated at some discrete-time index $N \ll 576 \cdot 10^3$. Secondly, the truncated RIRs are set to start at the, in absolute value, largest value of the original RIR recording, i.e. at the point where the direct component hits the microphone.

C. Hyperparameter Tuning

As with any algorithm requiring hyperparameters, the parameters of the proposed algorithm need to be rigorously set, in order for the algorithm to be effective. In this Section we expand on the tuning of the hyperparameters for the compared methods. The dependance on the variables of the problem largely decouples between the hyperparameters, with the optimal choice of γ being impacted by the input signal $\mathbf{x}$, and λ_{QE}, λ_{NN}, and λ_{T} depending on the SNR, much like in sparse estimation. Using a random subset of 20 of the RIRs of SMARD, we here vary the $\mathrm{SNR}_{\mathrm{dB}}$ for white Gaussian measurement noise from 0 to 20 in steps of 2, and find the optimal choice of λ_{QE}, λ_{NN}, and λ_{T} respectively, for each value of the SNR. We then perform linear regression to find models for λ_{QE}, λ_{NN}, and λ_{T}, as a function of the SNR. This is done with $n_h = 3600$, $n_y = 4320$, $\gamma = 10^2$ and with white Gaussian noise of unit power as input. In accordance with the results of this, we let $\lambda_{\mathrm{QE}} = 10^{0.4-0.14\cdot\mathrm{SNR}_{\mathrm{dB}}}$, $\lambda_{\mathrm{NN}} = 10^{-3.7-0.08\cdot\mathrm{SNR}_{\mathrm{dB}}}$, and $\lambda_{\mathrm{T}} = 10^{3.6-0.1\cdot\mathrm{SNR}_{\mathrm{dB}}}$. For λ_{QE} and λ_{NN}, this can is illustrated in Fig. 3. Different choices of n_h and n_y would yield different values for the model parameters, hence simulations from hereon throughout this paper will be done with $n_h = 3600$ and $n_y = 4320$, respectively. The performance of the proposed algorithms, as a function of SNR, will be discussed in Section V-D.

As for γ, the main dependency is on the input $\mathbf{x}$, but the dependency is more involved. The optimal choice is impacted by the power of the input signal, as well as the length of the RIR, n_h (and thereby the length of the input signal, n_x), and finally also the frequency content of the input signal. It is therefore very difficult to find a model that covers all possible choices of these variables. For a vast majority of the experiments performed, preliminary simulations showed that $\gamma = 10^2$ was a good enough choice. In Sections V-G and V-I, where the frequency content of the input signal is varied, each scenario requires its own opitimized γ, which can be found in Table I. In Section V-H, where snippets of speech are used as input, $\gamma = 10^{-2}$ proved to be a good choice.

D. SNR

The benefit of low-rank modeling of the RIR becomes particularly evident when the signal-to-noise ratio is low. With white Gaussian noise as input and values of λ_{QE}, λ_{NN}, λ_{T}, and γ as described in Section V-C, and using all of the 1008 RIRs from SMARD, we showcase the adequacy of the proposed modeling framework. The results can be seen in Fig. 4. The low-rank enforcing methods are clearly preferable to ordinary least squares, as well as Tikhonov regularized least squares. As expected, the difference compared to ordinary least squares is particularly apparent when the SNR is low.

E. Convergence Analysis

For the least squares methods, ordinary and with Tikhonov regularization, we have closed-form expressions, and no iter-

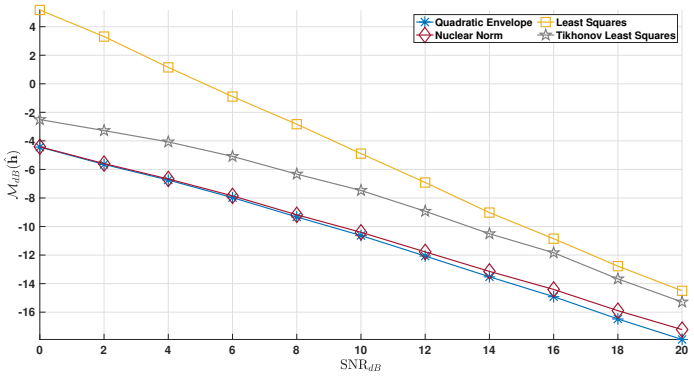

Fig. 4. Normalized misalignment as a function of $\mathrm{SNR}_{\mathrm{dB}}$.

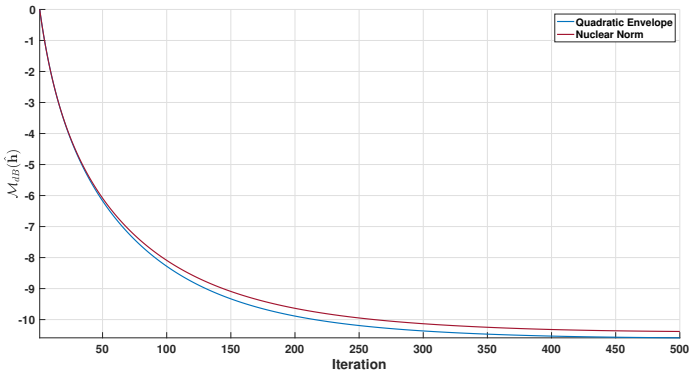

Fig. 5. Normalized misalignment as a function of number of iterations.

ative procedure is needed. For the nuclear norm and quadratic envelope, however, we here showcase the difference in speed of convergence, using part of the simulations done in Section V-D. In Fig. 5, we see the averaged normalized misalignment for the two iterative methods, as a function of the iteration index, for the case where $\mathrm{SNR}_{\mathrm{dB}} = 10$. The difference in average number of iterations needed for convergence for the quadratic envelope and nuclear norm, for this particular value of SNR, was negligible. We see in Fig. 5 that the normalized misalignment, for a given number of iterations, is lesser for the quadratic envelope, compared to the nuclear norm. Preliminary simulations showed similar behaviour for other SNR values, both in terms of the relation between the number of iterations needed for convergence of the respective methods, and the decay of the normalized misalignment, as a function of the iteration index.

F. Hyperparameter Sensitivity

In Section V-D we saw that the quadratic envelope and nuclear norm regularizations performed similarly as a function of the SNR, when oracle knowledge of the SNR was assumed. Using all the 1008 RIRs of SMARD, with $\mathrm{SNR}_{\mathrm{dB}} = 10$, and white Gaussian noise as input, we here consider what happens if the algorithms are provided over- or underestimated values of their respective regularization parameters. According to the formulas provided in Section V-C, we set $\lambda_{\mathrm{QE}} = 10^{0.4-0.14\cdot 10}$, $\lambda_{\mathrm{NN}} = 10^{-3.7-0.08\cdot 10}$, and $\lambda_{\mathrm{T}} = 10^{3.6-0.1\cdot 10}$ respectively. We then look at what happens if these parameters are multiplied by

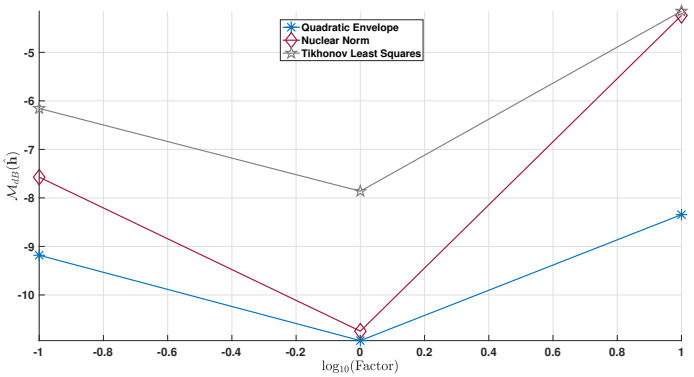

Fig. 6. Normalized misalignment with perturbed parameters.

TABLE I
PARAMETER VALUES USED WHEN VARYING α.

α	γ	λ_{QE}	λ_{NN}	λ_{T}
-1	$10^{-4.2}$	10^{4}	$10^{-5.4}$	10^{2}
-0.5	$10^{-3.8}$	10^{4}	10^{-5}	$10^{2.2}$
0	$10^{-3.2}$	10^{4}	$10^{-4.6}$	$10^{2.6}$
0.5	$10^{-2.4}$	10^{4}	$10^{-4.8}$	10^{3}
1	$10^{-1.8}$	10^{4}	$10^{-5.6}$	$10^{3.6}$

a factor of 10^{-1} or 10^{1}, so as to simulate an underestimation and an overestimation respectively. The averaged results are found in Fig. 6. There it can be seen that the quadratic envelope clearly is the method losing the least in accuracy, for the perturbed parameter values. This suggests that the quadratic envelope would be particularly useful in scenarios where the SNR is not known exactly. For reference, the average misalignment using ordinary least squares is on the order of -5 dB, indicating that when the knowledge of the SNR is poor and the regularization parameter is overestimated, some types of regularization do more harm than good.

G. Frequency Content of Input Signal

In this Section, we investigate how the proposed method performs as a function of inverse frequency power of the input, i.e. colored noise with the power spectral density $1/|f|^{\alpha}$. As previously mentioned, the hyperparameter values' dependency on the variables of the problem is complicated. Therefore, when displaying the aptitude of the proposed method as a function of α, the hyperparameters were tuned on 20 randomly selected RIRs. These parameter values are shown in Table I. Then, with these hyperparameter values, simulations were done on all of the RIRs of SMARD, and with $\mathrm{SNR}_{\mathrm{dB}} = 10$. The results are displayed in Fig. 7. There it can be see that the low-rank estimation methods, nuclear norm regularization as well as the quadratic envelope of the ℓ_0-norm, are particularly successful, in comparison to ordinary least squares, when the color of the noise is not white. As for the performance of Tikhonov least squares, we note that it is fairly constant as a function of α. Thus, the difference in performance between the low-rank methods and Tikhonov least squares is largest for α close to zero.

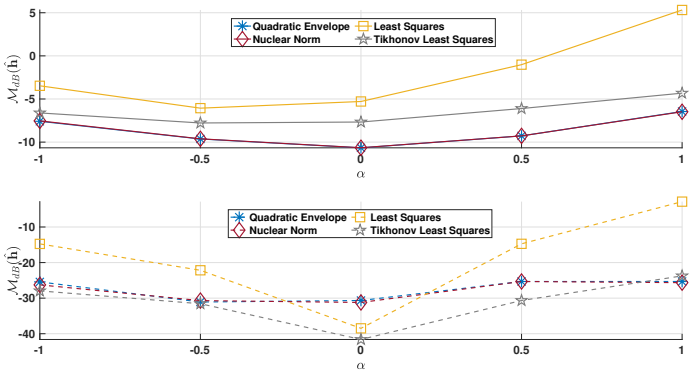

Fig. 7. Normalized misalignment as a function of α. Average (top) and standard deviation (bottom), respectively.

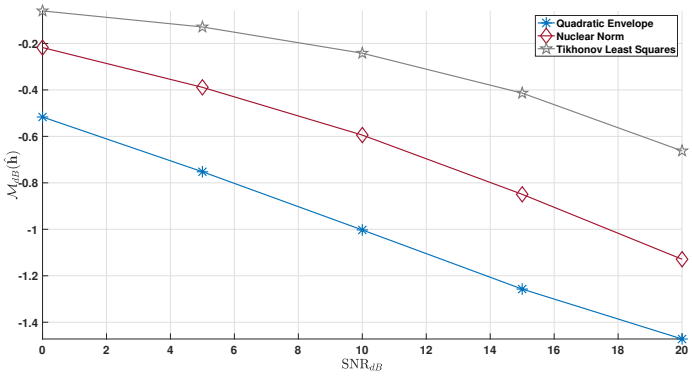

Fig. 8. Average normalized misalignment, using speech signals, as a function of SNR_{dB}.

H. Speech Signal Input

As indicated in Section V-G, the proposed method provides the most improvement to the RIR estimation when the spectral excitation by the input signal is poor. We study this input by letting the driving signal be a speech signal. Yet again, we set all the parameter values in accordance with the discussion in Section V-C (note that here $\gamma = 10^{-2}$), and use 100 randomly chosen RIRs from SMARD, all driven by 10 different snippets of anechoic male and female speech, chosen at random from the TSP Speech Database [54]. In order to make sure that the speech snippets contained enough power, they where all chosen with the criterion that $\|\mathbf{x}\| \geq 2$. The results can be seen in Fig. 8. Here, we have opted to not plot the normalized misalignment of the ordinary least squares estimation. It was consistently at least two orders of magnitude larger than the other methods, and the inclusion of it in Fig. 8 would cloud the difference between the other methods. It can be seen that, the low-rank methods outperform Tikhonov regularized least squares. This further strengthens the hypothesis that the low-rank modeling is most useful when the input spectral excitation is poor.

I. Input-Output Relations

In order to further investigate the merit of the low-rank estimation, we will in this Section, as previously indicated, use a slightly modified evaluation measure. Instead of, as previously, looking at the distance between the estimated RIR and the measured RIR, we will here look at the following metric,

$$\mathcal{M}^{O}_{\text{dB}}\left(\hat{\mathbf{h}}\right) = 20\log_{10}\left(\mathbb{E}\left[\frac{\|\hat{\mathbf{h}} * \mathbf{x} - \mathbf{h} * \mathbf{x}\|_2}{\|\mathbf{h} * \mathbf{x}\|_2}\right]\right), \quad (31)$$

where $\mathbf{x}$ is a previously unseen signal, with the same properties as the input signal generating the output $\mathbf{y}$, from which $\hat{\mathbf{h}}$ is estimated, and the expectation is with respect to $\mathbf{x}$. We will refer to this as *normalized output misalignment*. Under these conditions, we replicate the experiment in Section V-G. The same parameter values as in Section V-G are used, i.e. the ones found in Table I. Here, a subset of 100 randomly chosen RIRs from SMARD were used, and (31) was estimated using 50 different instances of $\mathbf{x}$, drawn from the same distribution as the input signal. The results can be seen in Fig. 9. It can be noticed that, for all but ordinary least squares, the normalized output misalignment is decreasing as a function of α. One possible explanation for this is the discrepancy in spectral excitation. For $\alpha = -1$, poor estimation at a certain frequency is more likely to be masked by poor spectral excitation by the input signal, at that particular frequency. The benefit of the low-rank regularization is, however, evident all across the board.

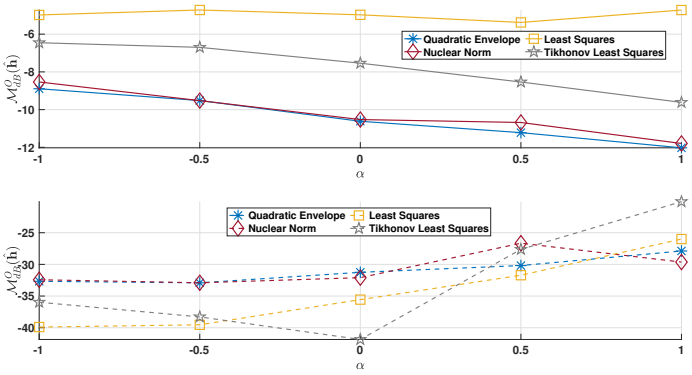

Fig. 9. Normalized output misalignment as a function of α, different evaluation signal. Average (top) and standard deviation (bottom), respectively.

J. Cramér–Rao Lower Bound

In this Section we study the mean squared error (MSE),

$$\text{MSE}\left(\hat{\mathbf{h}}\right) = \mathbb{E}\left[\|\hat{\mathbf{h}} - \mathbf{h}\|_2^2\right], \quad (32)$$

of the suggested estimators, in relation to theoretical bounds given by the Cramér–Rao Lower Bound (CRLB), the smallest variance that can be achieved by an unbiased estimator. Let

$$\mathbf{\Sigma}_{\mathbf{h}} = \mathbb{E}\left\{\left(\hat{\mathbf{h}} - \mathbf{h}\right)\left(\hat{\mathbf{h}} - \mathbf{h}\right)^T\right\} \quad (33)$$

denote the error covariance matrix of an unbiased estimator $\hat{\mathbf{h}}$. We have that $\text{MSE}\left(\hat{\mathbf{h}}\right) = \text{tr}(\mathbf{\Sigma}_{\mathbf{h}})$, and it can also be noticed that

$$\text{MSE}\left(\hat{\mathbf{h}}\right) = \text{Var}\left(\hat{\mathbf{h}}\right) + \text{Bias}\left(\hat{\mathbf{h}}, \mathbf{h}\right)^2. \quad (34)$$

We then define

$$\mathrm{MSE}_{\mathrm{dB}}\left(\hat{\mathbf{h}}\right) = 20\log_{10}\left(\mathrm{MSE}\left(\hat{\mathbf{h}}\right)\right). \tag{35}$$

The CRLB of the variance of low-rank estimation algorithms has been explored by Tang *et. al* in [55]. For the convenience of the reader, we here reiterate the principal content of their results. Let $\mathbf{\Sigma}$ denote the measurement noise covariance matrix. We then have that

$$\mathbf{\Sigma_h} \geq \mathbf{P}[\mathbf{P}^T\left(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right)\mathbf{P}]^{-1}\mathbf{P}^T, \tag{36}$$

with

$$\mathbf{P} = \begin{bmatrix}\bar{\mathbf{V}}\otimes\tilde{\mathbf{U}} & \tilde{\mathbf{V}}\otimes\tilde{\mathbf{U}} & \tilde{\mathbf{V}}\otimes\bar{\mathbf{U}}\end{bmatrix} \in \mathbb{R}^{q^2\times 2qr-r^2}, \tag{37}$$

$\mathbf{U} = \begin{bmatrix}\tilde{\mathbf{U}} & \bar{\mathbf{U}}\end{bmatrix} = [\mathbf{U}_1\,\mathbf{U}_2\,\ldots\mathbf{U}_r \quad \mathbf{U}_{r+1}\ldots\mathbf{U}_n]$, $\mathbf{V} = \begin{bmatrix}\tilde{\mathbf{V}} & \bar{\mathbf{V}}\end{bmatrix} = [\mathbf{V}_1\,\mathbf{V}_2\,\ldots\mathbf{V}_r \quad \mathbf{V}_{r+1}\ldots\mathbf{V}_p]$, $\mathbf{U}_j$ and $\mathbf{V}_j$ denoting the jth column of $\mathbf{U}$ and $\mathbf{V}$ respectively, r is the rank, and we remind the reader that $\mathbf{U}$ and $\mathbf{V}$ come from the SVD of $\mathbf{H} = \mathbf{USV}^T$, and that $\mathbf{X}$ is the Toeplitz matrix from (3). Further, $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. It follows from (36) that, by taking the trace of both sides of the inequality, using the cyclic property of the trace operation, and the fact that $\mathbf{P}^T\mathbf{P} = \mathbf{I}_{(2q-r)r}$,

$$\mathrm{MSE} \geq \mathrm{tr}\left(\left[\mathbf{P}^T\left(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right)\mathbf{P}\right]^{-1}\right), \tag{38}$$

i.e. the right hand side of (38) is the CRLB.

As previously touched upon, an RIR from SMARD will, when matricized, not be low-rank in a strict sense. In order to avoid complicated error bounds depending on the misspecification of the model, we will therefore, in this Section, make low-rank approximations of the matricized RIRs, in order to be able to study the properties of the estimation, in relation to the CRLB. Hence, $\mathbf{h}$ will in this Section be a rank-20 approximation of the recorded and truncated RIR. This is a meritable approximation, as per the discussion in [21]. A rigorous analysis of performance bounds for a not strictly low-rank RIR that has (purposefully) been misspecified as low-rank, is beyond the scope of this paper, and will be left for future work.

Comparison is made with two different theoretical bounds. The first one is under the assumption that the low-rank property is unknown. For the second bound, the low-rank property is known. Yet again, the parameter values from Section V-C are used. Experiments were performed using 50 randomly chosen RIRs from SMARD, and for each RIR running 20 Monte Carlo-simulations. The results, averaged over the Monte Carlo-simulations and the different RIRs, can be seen in Fig. 10. As expected, the first bound is greater than the second one. The low-rank methods constantly outperform the first bound, which is to be expected, since they are using information not available to that theoretical bound. In relation to the second bound, the low-rank methods outperform the bound for lower values of SNR. This is possible since the estimator is biased. Also this outperformance is appealing to intuition. For a lower SNR, the gain that comes from the regularization in terms of improved conditioning, can reduce variance so that it outweighs the added bias, something that is generally not possible with an increasing SNR.

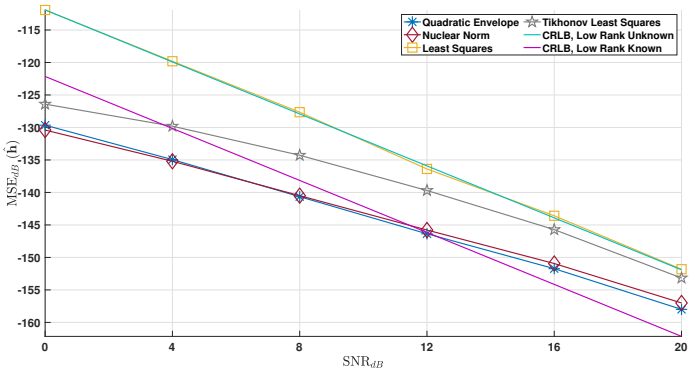

Fig. 10. Comparison with CRLB

VI. Conclusion

In this paper we have presented a method for exploiting low-rank, or close to low-rank, properties of RIRs. The low-rank penalty serves as a regularizer and prevents an overfitted solution, similar to the idea of sparse estimation. The low-rank penalty is enforced on a matricization of the coefficient vector, and consists of either the nuclear norm, or the quadratic envelope of the ℓ_0-pseudo-norm of the vector of singular values of the matrix. The results of the proposed methods are compared to those of ordinary least squares, as well as Tikhonov regularized least squares. It is shown that the low-rank methods work very well in general, but excel particularly when either the SNR is low, or when the input spectral excitation is poor, which can stem from the input signal being either colored noise or a speech signal. The difference between the two investigated low-rank methods, i.e. nuclear norm and quadratic envelope, is comparatively small, but with quadratic envelope having the benefit of superior robustness to a poorly chosen regularization parameter, in the case where oracle knowledge about the SNR can not be assumed.

Future work will focus on investigating the properties of the error introduced by the low-rank estimation. A structure to this error could indicate possible improvements, as well as potential applications, for the methods proposed here. The interplay between the bias introduced by a misspecification, and the possible reduced estimation variance, is another research area we aim to explore.

Appendix A
Quadratic Envelope of $\lambda\|x\|_0$

When trying to find $\mathcal{Q}_\gamma(\lambda_{\mathrm{QE}}\|\cdot\|_0)(u)$, we notice two things. Firstly, $\mathcal{Q}_\gamma(\lambda_{\mathrm{QE}}\|\cdot\|_0)(0) = 0$ and secondly, $\forall u$ s.t. $|u| \geq C$, for some constant $C > 0$, $\mathcal{Q}_\gamma(\lambda_{\mathrm{QE}}\|\cdot\|_0)(u) = \lambda_{\mathrm{QE}}$. The question is then what the value of C is, and what values $\mathcal{Q}_\gamma(\lambda_{\mathrm{QE}}\|\cdot\|_0)(u)$ takes on for $u \in (-C, C)$. The limiting factor for the envelope is that $\lambda_{\mathrm{QE}}\|0\|_0 = 0$, i.e. the negative quadratic function must pass through, or below, the origin. This yields the constraint that $\alpha - \frac{\gamma}{2}v^2 \leq 0 \Leftrightarrow \alpha \leq \frac{\gamma}{2}v^2$. Looking for the supremum, we want to chose the largest possible α, hence $\alpha = \frac{\gamma}{2}v^2$. Inserting this into (16) and simplifying yields

$$\sup_{v\in\mathbb{R}}\left\{\gamma uv - \frac{\gamma u^2}{2}; \gamma zv - \frac{\gamma z^2}{2} \leq \lambda_{\mathrm{QE}}\|z\|_0,\ \forall\, z\in\mathbb{R}\right\}. \tag{39}$$

We know this is valid in $z = 0$ so we can write

$$\sup_{v\in\mathbb{R}}\left\{\gamma uv - \frac{\gamma u^2}{2} : \gamma zv - \frac{\gamma z^2}{2} \leq \lambda_{\text{QE}},\ \forall\, z \in \mathbb{R}\setminus\{0\}\right\}. \tag{40}$$

If the peak of the parabola is at or below $v = \lambda_{\text{QE}}$ then, naturally, all other points on the parabola will be as well. Therefore we insert $z = v$ and simplify

$$\sup_{y\in\mathbb{R}}\left\{\gamma uv - \frac{\gamma u^2}{2} : \frac{\gamma v^2}{2} \leq \lambda_{\text{QE}}\right\} = \sup_{v\in\mathbb{R}}\left\{\gamma uv - \frac{\gamma u^2}{2} : |v| \leq \mu\right\}, \tag{41}$$

where $\mu = \sqrt{2\lambda_{\text{QE}}/\gamma}$. The expression we aim to maximize, $\gamma uv - \frac{\gamma u^2}{2}$, is growing in v if $v > 0$, and in $-v$ if $v < 0$. In other words, we want to chose v as big as possible, in modulo, without ever overestimating $g(z)$, yielding $y = \text{sign}(u)\mu$. Inserting this value for v into $\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(u)$ and setting this expression equal to λ_{QE} yields that this occurs at $u = \pm\mu$, thus, the sought after point $C = \mu$. Taken together we have

$$\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(u) = \begin{cases}\sqrt{2\lambda_{\text{QE}}\gamma}u - \frac{\gamma}{2}u^2 & ,\ u < \mu\\ \lambda_{\text{QE}} & ,\ \mu \leq u\end{cases} \tag{42}$$

or, equivalently,

$$\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(u) = \lambda_{\text{QE}} - \frac{\gamma}{2}\max(\mu - u, 0)^2. \tag{43}$$

APPENDIX B
PROXIMAL OPERATORS

The regularization terms of (15) and (20) are both orthogonally invariant [46]. For the proximal operator of an orthogonally invariant function $g : \mathbb{R}^{q\times p} \to \mathbb{R}$, it holds that

$$\text{prox}_{\rho g}(\mathbf{H}) = \mathbf{U}\,\mathcal{D}(\text{prox}_{\rho g}(\sigma(\mathbf{H})))\,\mathbf{V}^T, \tag{44}$$

i.e., the proximal step is performed on the vector of singular values [46]. For (15), the proximal operator of the penalty term is given by element-wise applying the so-called soft-thresholding operator $\mathcal{S}_{\lambda_{\text{NN}}} : \mathbb{R}^q \to \mathbb{R}^q$ [56],

$$\mathcal{S}_{\lambda_{\text{NN}}}(\sigma(\mathbf{H})_n) = \max(\sigma(\mathbf{H})_n - \lambda_{\text{NN}}, 0), \tag{45}$$

$n = 1, 2, \ldots, q$ where $\sigma(\mathbf{H})_n$, denotes the nth element of $\sigma(\mathbf{H})$.

For the penalty term in (20), we make use of the fact that the function decouples and that we therefore, without loss of generality, can consider the one-dimensional case. Further, since this is to be used for singular values, we consider only $u \geq 0$ as argument and $z \geq 0$ as function value. Finally, in order for the proximal operator to be single-valued, we must have that $\gamma\rho < 1$, since this guarantees that the curvature is strictly positive everywhere.

The proximal operator, (21), consists of two terms, the original function $\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(z)$, and the proximity term, $\frac{1}{2\rho}(u - z)^2$. What is sought after is the argument z that minimizes the sum of these, for a given u. First, it can be concluded that for $u = 0$, where the proximal operator is not differentiable, the optimal choice of z is 0. This renders a cost of 0, so no choice can be better. Next, the question is if there is an $u > 0$, for which choosing a $z > 0$ yields a lower cost than $z = 0$ and, if so, what u? For $z > 0$ the proximal operator is differentiable. We can set that derivative equal to zero to find the optimal choice of z there, z_{opt}. This yields that

$$z_{\text{opt}} = \begin{cases}\frac{u-\mu\gamma\rho}{1-\gamma\rho} & ,0 < u < \mu\\ u & ,\mu \leq u.\end{cases} \tag{46}$$

Comparing the costs and determining when it is better to choose $z > 0$, as opposed to $z = 0$, is equivalent of finding for which x

$$\lambda_{\text{QE}} - \frac{\gamma}{2}\max(\mu - z_{\text{opt}}, 0)^2 + \frac{1}{2\rho}(u - z_{\text{opt}})^2 < \frac{1}{2\rho}u^2. \tag{47}$$

Inserting z_{opt} from (46) yields that $x > \mu\gamma\rho$. Taken together,

$$\text{prox}_{\rho\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)}(u) = \begin{cases}0 & ,0 \leq u \leq \mu\gamma\rho\\ \frac{u-\mu\gamma\rho}{1-\gamma\rho} & ,\mu\gamma\rho < u < \mu\\ u & ,\mu \leq u\end{cases}. \tag{48}$$

To the best of the authors' knowledge, the explicit expression has not been published before.

REFERENCES

[1] C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge: Acoustic source localization and tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1620–1643, 2020.

[2] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*. Springer, 2010.

[3] C. Schissler, P. Stirling, and R. Mehra, "Efficient construction of the spatial room impulse response," in *2017 IEEE Virtual Reality (VR)*, 2017, pp. 122–130.

[4] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[5] G. W. Elko, E. Diethorn, and T. Gaensler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC '03)*, Kyoto, Japan, Sep. 2003.

[6] J. Mourjopoulos and M. Paraskevas, "Pole and zero modeling of room transfer functions," *Journal of Sound and Vibration*, vol. 146, no. 2, pp. 281–302, 1991.

[7] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, 2017.

[8] C. Huszty, N. Bukuli, Á. Torma, and F. Augusztinovicz, "Effects of filtering of room impulse responses on room acoustics parameters by using different filter structures," *J. Acoust. Soc. Amer.*, vol. 123, p. 3617, 2008.

[9] G. Vairetti, "Efficient parametric modeling, identification and equalization of room acoustics," Ph.D. dissertation, KU Leuven, 2018.

[10] L. S. H. Ngia, "Recursive identification of acoustic echo systems using orthonormal basis functions," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 278–293, 2003.

[11] K. Shi, X. Ma, and G. Tong Zhou, "An efficient acoustic echo cancellation design for systems with long room impulses and nonlinear loudspeakers," *Signal Processing*, vol. 89, no. 2, pp. 121–132, 2009.

[12] L. Krishnan, P. D. Teal, and T. Betlehem, "A robust sparse approach to acoustic impulse response shaping," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 738–742.

[13] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *International Journal of Control*, vol. 93, no. 2, pp. 173–180, 2020.

[14] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Processing*, vol. 88, no. 3, pp. 594–611, 2008.

[15] Z. Hu, F. Nie, R. Wang, and X. Li, "Low rank regularization: A review," *Neural Networks*, vol. 136, pp. 218–232, 2021.

[16] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" arXiv 0912.3599, 2009.

[17] Q. Yao, J. Kwok, T. Wang, and T. Liu, "Large-scale low-rank matrix learning with nonconvex regularizers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2628–2643, 2019.

[18] M. Boussé and L. De Lathauwer, "Large-scale autoregressive system identification using kronecker product equations," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 1348–1352.

[19] A. Marconato, L. Ljung, Y. Rolain, and J. Schoukens, "Linking regularization and low-rank approximation for impulse response modeling," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 4999–5004, 2014, 19th IFAC World Congress.

[20] A. Marconato and J. Schoukens, "Impact of regularization in fir estimation for short and long data records," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015, pp. 789–793.

[21] M. Jälmby, F. Elvander, and T. van Waterschoot, "Low-rank tensor modeling of room impulse responses," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 111–115.

[22] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, ser. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2012.

[23] C. Paleologu, J. Benesty, and S. Ciochină, "Linear system identification based on a Kronecker product decomposition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, 2018.

[24] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochină, "Recursive least-squares algorithms for the identification of low-rank systems," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 5, pp. 903–918, 2019.

[25] L.-M. Dogariu, C. Paleologu, J. Benesty, and S. Ciochină, "An efficient Kalman filter for the identification of low-rank systems," *Signal Processing*, vol. 166, p. 107239, 2020.

[26] Y. Xiang, S.and Zhu, X. Shen, and J. Ye, "Optimal exact least squares rank minimization," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 480–488.

[27] M. Carlsson, "On convex envelopes and regularization of non-convex functionals without moving global minima," *Journal of Optimization Theory and Applications*, vol. 183, no. 1, pp. 66–84, 2019.

[28] V. Larsson and C. Olsson, "Convex low rank approximation," *International Journal of Computer Vision*, vol. 120, no. 2, pp. 194–214, 2016.

[29] C. Olsson, M. Carlsson, F. Andersson, and V. Larsson, "Non-convex rank/sparsity regularization and local minima," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 332–340.

[30] J. Caprani and M. Carlsson, "Quadratic envelope regularization for structured low rank approximation," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8217–8221.

[31] P. C. Hansen, "Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank," *SIAM Journal on Scientific and Statistical Computing*, vol. 11, no. 3, pp. 503–518, 1990.

[32] N. Antonello, L. Stella, P. Patrinos, and T. van Waterschoot, "Proximal gradient algorithms: Applications in signal processing," arXiv: 1803.01621, 2020.

[33] H. Kuttruff, *Room Acoustics*. London: Spon Press, 2009.

[34] F. Jacobsen and P. M. Juhl, *Fundamentals of General Linear Acoustics*. Wiley, 2013.

[35] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, 2017.

[36] L. D. Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-(L_r,L_r,1) terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1451–1474, 2011.

[37] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the 2001 American Control Conference*, vol. 6, 2001, pp. 4734–4739 vol.6.

[38] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[39] L. Gu, S.and Zhang and X. Zuo, W.and Feng, "Weighted nuclear norm minimization with application to image denoising," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[40] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, Feb. 2015.

[41] E. Soubies, A. Chinatto, P. Larzabal, J. Romano, and L. Blanc-Féraud, "Direction-of-arrival estimation through exact continuous $\ell_{2,0}$-norm relaxation," *IEEE Signal Processing Letters*, vol. 28, pp. 16–20, 2021.

[42] E. Vasquez-Ortiz and P. Rodriguez, "Fast gradient-based algorithm for a quadratic envelope relaxation of the ℓ_0 gradient regularization," in *2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, 2021, pp. 1–6.

[43] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, pp. 894 – 942, 2010.

[44] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A continuous exact ℓ_0 penalty (cel0) for least squares regularized problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1607–1639, 2015.

[45] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*. New York, NY: Springer New York, 2011, pp. 185–212.

[46] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[47] H. Attouch, J. Bolte, and B. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Mathematical Programming*, vol. 137, no. 1, pp. 91–129, 2013.

[48] L. Stella, A. Themelis, and P. Patrinos, "Forward–backward quasi-newton methods for nonsmooth optimization problems," *Computational Optimization and Applications*, vol. 67, no. 3, pp. 443–487, 2017.

[49] R. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[50] R. Badeau, "Unified stochastic reverberation modeling," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 2175–2179.

[51] M. R. Schroeder, "Frequency-correlation functions of frequency responses in rooms," *The Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1819–1823, 1962.

[52] J. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, pp. 605–639, 01 1985.

[53] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single- and multichannel audio recordings database (SMARD)," in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC '14)*, Antibes, France, Sep. 2014.

[54] P. Kabal, "TSP speech database," McGill University, Tech. Rep., 2002.

[55] G. Tang and A. Nehorai, "Lower bounds on the mean-squared error of low-rank matrix reconstruction," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4559–4571, 2011.

[56] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.



Martin Jälmby received the M.Sc. degree in engineering mathematics from Lund University, Sweden in 2018. He is currently working toward the Ph.D. degree in electrical engineering with KU Leuven, Belgium, where his research focuses on low-rank modeling in acoustics.

PLACE PHOTO HERE

Filip Elvander received the M.Sc. degree from industrial engineering and management and the Ph.D. in mathematical statistics from Lund University, Sweden in 2015 and 2020, respectively. He is currently a Postdoctoral Research Fellow with the Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Belgium. His research interests include inverse problems, robust estimation, and convex modeling and approximation techniques in statistical signal processing and spectral analysis.

PLACE PHOTO HERE

Toon van Waterschoot (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from KU Leuven, Leuven, Belgium, in 2001 and 2009, respectively. He has held teaching and research positions with the Delft University of Technology, Delft, The Netherlands, and University of Lugano, Lugano, Switzerland. He is currently an Associate Professor and Consolidator Grantee with the European Research Council, KU Leuven. His research interests include signal processing, machine learning, and numerical optimization, applied to acoustic signal enhancement, acoustic modeling, audio analysis, and audio reproduction. He was the General Chair of the 60th AES International Conference with Leuven, Belgium in 2016, and serving on the Organizing Committee of the European Conference on Computational Optimization (EUCCO 2016), IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2017), and 28th European Signal Processing Conference (EUSIPCO 2020). He is also an Associate Editor for the Journal of the Audio Engineering Society and EURASIP Journal on Audio, Music, and Speech Processing, and the Guest Editor of Elsevier Signal Processing. He is also the Director of the European Association for Signal Processing (EURASIP), a Member of the IEEE Audio and Acoustic Signal Processing Technical Committee, a Member of the EURASIP Special Area Team on Acoustic, Speech and Music Signal Processing, and a FoundingMember of the EAA Technical Committee in Audio Signal Processing. He is a Member of the EURASIP, ASA, and AES.